Supporting Figures

Kinetic on-rate:

$$[C] = [L]_0 \, (1 - e^{-k_{on} \times [T]_0 \times t})$$

Kinetic off-rate:

$$[C] = [C]_0 \, e^{-k_{off} \times t}$$

Figure S1. On- and off-rate equations for the HTSK experiments. Due to the relatively short time period for the on-rate segment of the experiment (~45 minutes) and the very slow off-rate for the clones (~2 x 10^{-6} on average) the contribution from the off-rate can be ignored during the binding phase. This allowed the transient complex concentration equation under excess target concentration conditions to reduce to the kinetic on-rate expression above. To fit the HTSK data to the above model, the %C bound as a function of equilibrium value was obtained by dividing [C] by $[L]_0$. This allowed us to fit the data for 2 parameters $\%C_{max}$ and k_{on}. The kinetic off-rate was obtained by blocking the on-rate contribution to the transient binding model. The k_{off} value was obtained by fitting the HTSK data for 2 parameters: $\%C_{max}$ and k_{off}.

Sequence Frequency in Pool (Raw Data)							
Name	Sequence	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6
E5	MDFITIYNYKKAADHFSMSMGSGSGS	101	103	120	114	109	89
E4	MNLTIAIYNYKKAADHFSMSMGSGSGS	2,804	3,157	3,458	3,670	4,418	5,245
E6	MNTITIYNYKKAADHFPMSMGSGSGS	93	94	81	58	41	40
E7	MNLTTIYNYKEAADSFSLVMGSGSGS	743	757	551	319	231	354
E3	MIRIITIYNYKKAADHYAQVLGSGSGS	18	23	24	27	31	42
E8	MRTITIYNYKKAAGHYAQVLGSGSGS	221	248	241	216	193	185
E9	MMPTITIYNYKKAADHYAQVLGSGSGS	51	56	46	49	46	42
# of Sequences in the Pool		3,159,350	3,330,810	3,331,534	3,019,767	3,056,863	2,903,743

Sequence Fractional Composition (Parts Per Million)							
Name	Sequence	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6
E5	MDFITIYNYKKAADHFSMSMGSGSGS	32.0	30.9	36.0	37.8	35.7	30.7
E4	MNLTIAIYNYKKAADHFSMSMGSGSGS	888	948	1,038	1,215	1,445	1,806
E6	MNTITIYNYKKAADHFPMSMGSGSGS	29.4	28.2	24.3	19.2	13.4	13.8
E7	MNLTTIYNYKEAADSFSLVMGSGSGS	235	227	165	106	76	122
E3	MIRIITIYNYKKAADHYAQVLGSGSGS	5.7	6.9	7.2	8.9	10.1	14.5
E8	MRTITIYNYKKAAGHYAQVLGSGSGS	70.0	74.5	72.3	71.5	63.1	63.7
E9	MMPTITIYNYKKAADHYAQVLGSGSGS	16.1	16.8	13.8	16.2	15.0	14.5

Radiolabeled Counts Bound to Beads x 10^{-3} (Pool Behavior)						
Name	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6
Extension Pool	24.175	23.006	22.634	17.437	14.599	9.962

Amount of Peptide Remaining on Beads (PPM x Radiolabeled Counts)								Fitting Parameters	
Name	Sequence	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6	Maximum	Off-rate
E5	MDFITIYNYKKAADHFSMSMGSGSGS	773	711	815	658	521	305	780	2.2E-06
E4	MNLTIAIYNYKKAADHFSMSMGSGSGS	21,456	21,805	23,493	21,192	21,100	17,994	22028	4.8E-07
E6	MNTITIYNYKKAADHFPMSMGSGSGS	712	649	550	335	196	137	698	7.6E-06
E7	MNLTTIYNYKEAADSFSLVMGSGSGS	5,685	5,229	3,743	1,842	1,103	1,214	5851	1.4E-05
E3	MIRIITIYNYKKAADHYAQVLGSGSGS	138	159	163	156	148	144	159	2.3E-07
E8	MRTITIYNYKKAAGHYAQVLGSGSGS	1,691	1,713	1,637	1,247	922	635	1710	3.7E-06
E9	MMPTITIYNYKKAADHYAQVLGSGSGS	390	387	313	283	220	144	395	3.5E-06
Time (s)		1,980	9,000	30,540	80,340	171,600	428,940		

% Counts Remaining on Beads (Normalized to Maximum)							
Name	Sequence	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5	Time Point 6
E5	MDFITIYNYKKAADHFSMSMGSGSGS	99%	91%	105%	84%	67%	39%
E4	MNLTIAIYNYKKAADHFSMSMGSGSGS	97%	99%	107%	96%	96%	82%
E6	MNTITIYNYKKAADHFPMSMGSGSGS	102%	93%	79%	48%	28%	20%
E7	MNLTTIYNYKEAADSFSLVMGSGSGS	97%	89%	64%	31%	19%	21%
E3	MIRIITIYNYKKAADHYAQVLGSGSGS	87%	100%	103%	98%	93%	91%
E8	MRTITIYNYKKAAGHYAQVLGSGSGS	99%	100%	96%	73%	54%	37%
E9	MMPTITIYNYKKAADHYAQVLGSGSGS	99%	98%	79%	72%	56%	37%
Time (s)		1,980	9,000	30,540	80,340	171,600	428,940

Table S1. Steps for obtaining the HTSK dissociation rates. Each time point during dissociation was sent for high-throughput sequencing. Top table: The frequency of each sequence at each time point. The total number of sequences in each pool is in the bottom row. Second table from the top: Diving the sequence frequency by the total number of sequence in the pool provides the fraction of pool comprised of each sequence (fractional composition). We express this number in terms of parts per million (PPM). Middle table: The radiolabeled counts bound to the beads at each time point. Second table from the bottom: Multiplying the radiolabeled counts bound to beads by the composition (PPM values) gives us the amount of each peptide bound to beads at each time point. These values can now be used to obtain the off-rates by fitting. Bottom table: by dividing the values by the asymptotic maximum, obtained during fitting, we are able to obtain the % Counts remaining on the beads for each peptide. This is the data shown in Figure 1c, right panel.

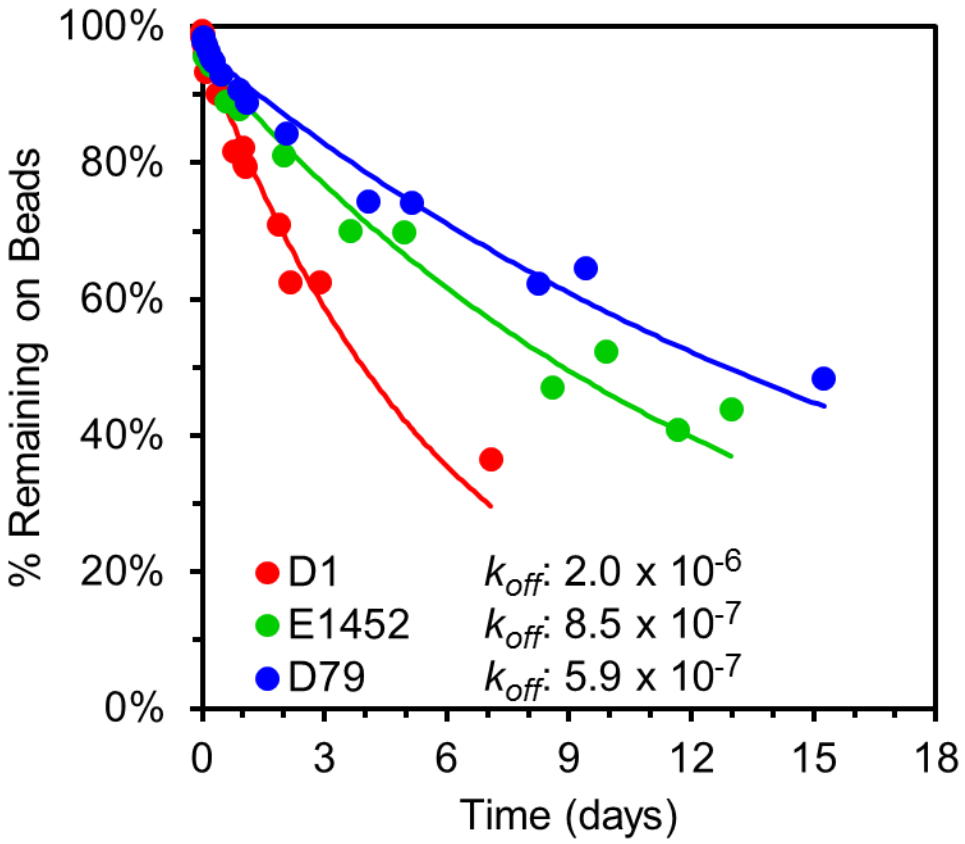

Figure S2. Ligand E1452 (green circles, frequency rank of 1452 in the extension selection pool) was identified by HTSK and tested as a radiolabeled peptide. Its off-rate is slower than D1, the previously identified highest affinity peptide from the doped selection. These results point to the ability of the extension selection to generate ultra-high affinity ligands without the need for a biased (doped) selection to improve affinity further.

	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})
Radiolabeled	1.0×10^{5}	2.9×10^{-6}
DNA Quantitation	1.3×10^{5}	2.4×10^{-6}

Table S2. Non-radiolabeled methods of obtaining pool behavior. It is possible to obtain the on- and off-rates for the pool by quantitation of the DNA bound to beads at each time point. The table above shows the obtained kinetic rates for the extension pool by DNA quantitation as well as radiolabeled binding.

	Sequence	Peptide k_{off} (s^{-1})	HTSK k_{off} (s^{-1})	ELISA K_d (pM)	HTSK K_d (pM)
E1	MIETITIYNYKKAADHFSMSM	7.4×10^{-6}	2.5×10^{-6}	39 ± 6*	23 ± 2
D1	--AIS----------YA-TK	2.0×10^{-6}	1.0×10^{-6}	9 ± 2*	15
D79	--D-NV-L----------IT-	5.9×10^{-7}	3.3×10^{-7}		2.4

* Jalali-Yazdi *et al*.[1]

Table S3. Validity of the HTSK results. The kinetic off-rates and the dissociation constant for three selected clones obtained by HTSK vs. radiolabeled peptides (k_{off}) and ELISA (K_d).

Materials and Methods:

Protein Expression and Purification. The gene for the first 209 amino acids of Bcl-x_L (Clone HsCD00004711; Dana Farber/Harvard Cancer Center DNA Resource Core) was PCR amplified with Pfusion polymerase. An N-terminal avitag (AGGLNDIFEAQKIEWHEGG) was added via the PCR reaction for *in vivo* biotinylation using the BirA enzyme.[2] The product was purified via PCR purification column and cloned into the pET24a vector using NdeI and XhoI. Bcl-x_L was expressed overnight at 37 °C in BL21(DE3) cells using auto-induction media.[3] Cells were lysed using Bper (Pierce), and purified using Ni-NTA superflow resin on an FPLC (Bio-Rad), using a gradient from 10 mM to 400 mM imidazole (Buffer A: 25 mM Hepes pH 7.5, 1 M NaCl, 10 mM imidazole; Buffer B: 25 mM Hepes pH 7.5, 1 M NaCl, 400 mM imidazole). Fractions with pure Bcl-x_L were combined, concentrated, and desalted into 50 mM Tris-HCl, pH 8.0. Bcl-x_L was biotinylated *in vitro* using BirA biotin ligase (0.1 mg/mL in 50 mM Tris-HCl, pH 8.3, 10 mM ATP, 10 mM $Mg(OAc)_2$, 50 µM biotin) at 30 °C for two hours. The protein was buffer exchanged into 1X PBS, frozen in liquid nitrogen, and stored at -80 °C.

Peptide Synthesis. Peptides E1 (NH_2-MIETITIYNYKKAADHFSMSMGSK-NH_2), E2 (NH_2-MIETITIYKYKKAADHFSMSMGSK-NH_2), D1 (NH_2-MIAISTIYNYKKAADHYAMTKGSK-NH_2), and D79 (NH_2-MIDTNVILNYKKAADHFSITMGSK-NH_2) were synthesized by solid phase Fmoc synthesis, using a Biotage Alstra Microwave Synthesizer.[4] The peptides were synthesized on Rink amide MBHA resin using five-fold molar excess of each amino acid and HATU. After the coupling of the first amino acid, (Fmoc-Lys(Mtt)-OH), the primary amine in the side-chain of the lysine for each peptide was deprotected using a solution of 1% (v/v) trifluoroacetic acid (TFA) in Dichloromethane (DCM). Biotin was then coupled to the side-chain primary amine before the synthesis was resumed, resulting in biotin-labeled peptides. Peptides were cleaved from the resin and deprotection with a solution of 95% (v/v) TFA, 2.5% 1,2-ethanedithiol (EDT), 1.5% (v/v) deionized water (DI), and 1% (v/v) triisopropylsilane (TIS) for 2 hours at room temperature.[5] The resin was filtered out, and the peptide was precipitated using 4-fold (v/v) excess ether. The peptides were dried, resuspended in DMSO, and HPLC purified using a C_{18} reverse phase column and a gradient of 10-90% acetonitrile/0.1% TFA in water. Fractions were collected and tested for the correct molecular weight using MALDI-TOF mass spectrometry. The correct fractions were lyophilized, dissolved in DMSO, and flash frozen at -80 °C.

Radiolabeled Off-Rate Assay.
The DNA sequences coding for the peptides were ordered from Integrated DNA Technologies (IDT). Each DNA construct contained a T7 RNA Polymerase promoter, and a 5' deletion mutant of the Tobacco Mosaic Virus (ΔTMV).[6] The C-terminal portion of the peptides were elongated with a flexible serine-glycine linker (six amino acids long) and an HA tag. After gel purification using urea-PAGE, the DNA sequences were PCR amplified using Taq polymerase and *in vitro* transcribed into mRNA using T7 RNA

polymerase.[6] After transcription, the mRNA was urea-PAGE purified and resuspended in deionized water to a final concentration of 30 μM.

The samples were *in vitro* translated at 30 °C for 1 hour in the translation solution—150 mM KOAc, 750 μM $MgCl_2$, 2 μM mRNA, 1X translation mix (20 mM Hepes-KOH pH 7.6, 100 mM creatine phosphate, 2 mM DTT, and 312.5 μM of each amino acid excluding methionine), ^{35}S-labeled methionine (Perkin Elmer; 20 μCi for each 25 μL of translation), and 60% (v/v) rabbit reticulocyte lysate (Green Hectares; prepared according to the method of Jackson and Hunt)[7] . Radiolabeled peptides were purified using magnetic HA beads (Life Technologies) and eluted with 100 μL, 50 mM NaOH, then immediately neutralized with 20 μL of 1 M Tris-HCl, pH 8.0.

The radiolabeled peptides were allowed to bind to 30 pmol immobilized Bcl-x_L for 1 hour in sample buffer (1X PBS, 1% (w/v) BSA, 0.1% (v/v) Tween 20, 10 μM biotin). The beads were magnetically separated, and washed 5X with sample buffer. The beads were resuspended in 1 mL of sample buffer containing 3 μM non-biotinylated Bcl-x_L (~100X molar excess relative to immobilized biotinylated Bcl-x_L). At various time points, 100 μL of slurry was removed and the beads were magnetically separated and washed. The percent remaining at each time point was determined by dividing the counts per minute (cpm) on beads by total cpm (beads + washes). The peptide off-rate was determined by an exponential fit of the Percent counts on beads vs. Time (s).

Enzymatic K_d Calculation Assay. The K_d values of the peptides were determined using a protocol modified from Friguet et al.[8] A set of serially diluted Bcl-x_L standards, at 2X the desired concentration, were made in sample buffer. For each peptide ligand, a set of dilutions at 2X the desired concentrations were also prepared. The Bcl-x_L samples were either mixed 1:1 with sample buffer (standards) or ligands (samples), and allowed to incubate at room temperature for 6 days.

To analyze the samples, the ELISA plates were incubated overnight at 4 °C with 1.5 nmol of streptavidin in 1X PBS. Plates were washed 3X with wash buffer (1X PBS + 0.1% (v/v) Tween-20) and blocked with 1X PBS + 5% (w/v) Bsa for two hours. 100 □L of a 30 nM solution of the D1 peptide (capture ligand) was added to wells and incubated for 1 hour. After the capture ligand incubation, 100 □L of sample or standards were added in each well, and incubated for 1 hour at room temperature. Plates were washed, incubated with HRP-conjugated anti-HIS tag antibody in sample buffer for 1 hour, washed, and incubated with TMB substrate (Thermo Scientific). Reactions were stopped after approximately 10 minutes with 2 M sulfuric acid, and the absorbance at 450 nm was measured via a plate reader (Molecular Devices).

The OD450 for the standards and their concentration values were fit to a four parameter logistic curve (standard curve). The concentration of the free Bcl-x_L in solution (responsible for the signal) for each sample was calculated using the standard curve, and converted into percent of Bcl-x_L bound by peptide in solution. For each peptide, the values for all the tested concentration of Bcl-x_L and peptide in solution were fit simultaneously to the monovalent equilibrium model to obtain the K_d.

$$[C]_{EQ} = \frac{[T]_0 + [L]_0 + K_D - \sqrt{([T]_0 + [L]_0 + K_D)^2 - 4[T]_0[L]_0}}{2}$$

Preparing the Pools. The DNA for the final enriched pools from the extension and the doped selection against Bcl-x_L were obtained from Dr. Takahashi (Takahashi and Roberts, manuscript in preparation). The DNA pools were PCR amplified using Taq polymerase and *in vitro* transcribed into mRNA using T7 RNA polymerase.[6] After transcription, the mRNA was urea-PAGE purified and resuspended in deionized water to a final concentration of 30 μM. The mRNA was then ligated to fluorescein-F30P (phosphate–dA_{21}–[dT-fluor]–$[C9]_3$–dAdCdCP; where [dT-fluor] is fluorescein dT (Glen Research), [C9] is spacer 9 (Glen Research), and P is puromycin (Glen Research); synthesized at the Keck Oligo Facility at Yale) using T4 DNA ligase.[9] The ligation was performed using a splint complementary to the 3' end of the RNA and the 5' end of the DNA-linker. The ligated mRNA was urea-PAGE purified and resuspended in deionized water to final concentration of 30 μM. The samples were *in vitro* translated in the translation solution—150 mM KOAc, 750 μM $MgCl_2$, 2 μM mRNA, in 1X translation mix (20 mM Hepes-KOH pH 7.6, 100 mM creatine phosphate, 2 mM DTT, and 312.5 μM of each amino acid) and 60% (v/v) rabbit reticulocyte lysate (Green Hectares; prepared according to the method of Jackson and Hunt[7]). To prepare radiolabeled peptides or proteins, non-labeled methionine was substituted with ^{35}S-labeled methionine (Perkin Elmer; 20 μCi for each 25 μL of translation). The translation reactions were incubated at 30 °C for one hour. To form mRNA-protein fusions, KCl and $MgCl_2$ were added to the reaction to final concentrations of 250 mM and 30 mM respectively after translation, and the samples were frozen at -20 °C.

To purify the fusion molecules, 100 μL of dT cellulose (25% (v/v) slurry, GE Healthcare) in isolation buffer (100 mM Tris-HCl pH 8.0, 1 M NaCl, 0.2% (v/v) Triton X-100) was added and incubated for 1 hour. The beads were washed five times with 700 μL of isolation buffer, and the fusions were eluted with 3X 80 μL of 65 ºC water and desalted through Centrisep columns (Princeton Separations). The desalted fusions were adjusted to 1X RT buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 2.4 mM 3' primer, 200 mM each dNTP) and the sample was heated to 65 °C for 5 minutes and cooled on ice to anneal the 3' primer. After cooling, 10 μL of Superscript II enzyme was added and the reaction incubated at 42 °C for one hour. Superscript II was inactivated by heating to 65 °C for 5 minutes, after which the samples were cooled on ice, and used within the same day.

On- and off-rate experiments. To obtain high-throughput sequencing kinetic (HTSK) on-rates, mRNA-peptide fusions of each pool from a 50 μL translation reaction (radiolabeled and non-labeled fusions separately) were first mixed with 7.5 pmols of Bcl-x_L immobilized on magnetic beads, and adjusted to 1 mL in 1X Selection buffer (1X PBS, 0.1% (w/v) BSA, 0.1% (v/v) Tween20, 100 μg/mL yeast tRNA, 0.05% (w/v) sodium azide, 10 μM biotin). At each time point, 100 μL of the solution was removed. The non-radiolabeled samples were magnetically separated and washed, PCR amplified with the appropriate primers, and sent for next-generation sequencing. The radiolabeled samples were washed 3X, and the the beads were counted via a scintillation counter.

To obtain the HTSK off-rates, after the kinetic on-rate experiment, the remaining beads were washed 5X with selection buffer. The beads were then resuspended in 800 μL of selection buffer without biotin and supplemented with 2 μM Bcl-x_L in solution. The

excess Bcl-x_L in solution prevents binding of dissociated ligands back to the beads. At specific time points, 100 µL of the solution was removed. The non-radiolabeled samples were washed, PCR amplified, and sent for next-generation sequencing. The radiolabeled samples were washed and counted via a scintillation counter.

Next Generation DNA Sequencing Analysis. The mRNA-peptide fusions from all of the time points and pools were PCR amplified using unique identifying barcodes, combined into a single sample and sent for high throughput DNA sequencing using a HiSeq 2500 machine at the USC genome core. The file containing the results from the DNA sequencing run (FASTQ format) was first stripped of all content except for the DNA sequences using python code developed in house. Then the file was split into separate files for each on- and off-rate time point based on the DNA bar code. Each DNA sequence in each file was then translated (only the region after the start codon until the 3' primer, using biopython and in house developed code) and the frequency of each translated sequence in the pool was calculated. Then, the fractional composition (frequency of the sequence divided by the total sequences in the pool) for each sequence was calculated. A separate file was created per selection to track the frequency composition for each sequence throughout the various time points. An example of this data can be seen in Figure 1a-b in the left panels.

Obtaining the on- and off-rates by HTSK. To obtain the on-rate for each sequence, the fractional composition for each sequence was multiplied by the radiolabeled counts for that pool's time point. This results in the radiolabeled counts per sequence as a function of time. In cases where the sequence diversity is low, or there is a great change in composition of the pool during the various time points, it might become necessary to normalize the radiolabeled counts by each time point's average number of methionine residues per sequence. This calculation is easy to do, as the values can be easily obtained by counting the total number of methionine obtained by high throughput sequencing, and dividing by the total number of sequences in the pool.

The radiolabeled counts per sequence (representing [C]), the concentration of immobilized Bcl-x_L on magnetic beads, and time in seconds were fit to the on-rate equation in supplementary figure 1 to obtain $[L]_0$ (asymptotic maximum) and k_{on} for each sequence. The fitting was done using the fminsearch function in MATLAB to minimize the error (Least Absolute Deviation method) between the real data and the model by changing $[L]_0$ and k_{on}. To obtain the off-rate, the same procedure was performed with the off-rate portion of the fraction composition data for each sequence. MATLAB was used to fit the product of the fractional composition and the radiolabeled pool counts at each time point, to the off-rate formula in Supplementary Figure 1.

To obtain the on- and off-rates for each sequence without using the radiolabeled data, it is possible to use another method to quantitating the amount of pool bound to the beads at each time point. We quantitated the amount of DNA bound to the beads by measuring the intensity of the DNA bands in the agarose gels using ImageJ's intensity measurement tool, and using the DNA ladder (NEB 100bp ladder) as our standards.

Number of sequences that this analysis can give reliable results for depends on diversity and the status of the library. We could only do the analysis for ligands with a statistically significant representation in a pool. For a pool that had converged to a large degree (extension pool), where the top 50 sequences accounted for ~78% of the pool, we were able to obtain HTSK results for approximately 2,000 sequences. However for a less converged pool (Doped) where the top 50 sequences accounted for ~3% of the pool, we were able to obtain HTSK results for ~20,000 sequences. The HTSK analysis cannot however provide kinetics constants for any sequences if the diversity of the pool was too high (where the highest represented sequence in the library accounted for less than 1 PPM of the library).

References

[1] F. Jalali-Yazdi, T. T. Takahashi, R. W. Roberts, *Analytical chemistry* **2015**, *87*, 11755-11762.

[2] D. Beckett, E. Kovaleva, P. J. Schatz, *Protein Sci* **1999**, *8*, 921-929.

[3] F. W. Studier, *Protein Expr Purif* **2005**, *41*, 207-234.

[4] G. B. Fields, R. L. Noble, *Int J Pept Protein Res* **1990**, *35*, 161-214.

[5] H. Huang, D. L. Rabenstein, *J Pept Res* **1999**, *53*, 548-553.

[6] R. Liu, J. E. Barrick, J. W. Szostak, R. W. Roberts, *Methods Enzymol* **2000**, *318*, 268-293.

[7] R. J. Jackson, T. Hunt, *Methods Enzymol* **1983**, *96*, 50-74.

[8] B. Friguet, A. F. Chaffotte, L. Djavadi-Ohaniance, M. E. Goldberg, *Journal of immunological methods* **1985**, *77*, 305-319.

[9] T. T. Takahashi, R. W. Roberts, *Methods Mol Biol* **2009**, *535*, 293-314.